DOCUMENT RESUME

ED 459 825 IR 058 361

AUTHOR Muir, Adrienne
TITLE Legal Deposit of Digital Publications: A Review of Research and Development Activity.
PUB DATE 2001-06-00
NOTE 10p.; In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (1st, Roanoke, Virginia, June 24-28, 2001). For entire proceedings, see IR 058 348. Funded by the Loughborough University Faculty Studentship scheme.
AVAILABLE FROM Association for Computing Machinery, 1515 Broadway, New York NY 10036. Tel: 800-342-6626 (Toll Free); Tel: 212-626-0500; e-mail: acmhelp@acm.org. For full text: http://www1.acm.org/pubs/contents/proceedings/dl/379437/.
PUB TYPE Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; Archives; *Depository Libraries; *Electronic Libraries; Information Policy; Preservation; Publications; Research and Development
IDENTIFIERS *Digital Preservation; Digital Technology

ABSTRACT

There is a global trend towards extending legal deposit to include digital publications in order to maintain comprehensive national archives. However, including digital publications in legal deposit regulation is not enough to ensure the long-term preservation of these publications. Concepts, principles and practices accepted and understood in the print environment, may have new meanings or no longer be appropriate in a networked environment. Mechanisms for identifying, selecting and depositing digital material either do not exist, or are inappropriate, for some kinds of digital publication. Work on developing digital preservation strategies is at an early stage. National and other deposit libraries are at the forefront of research and development in this area, often working in partnership with other libraries, publishers and technology vendors. Most work is of a technical nature. There is some work on developing policies and strategies for managing digital resources. However, not all management issues or users' needs are being addressed. This review of research and development work focuses on activities specifically related to digital legal deposit. It also touches on more generic work that is especially relevant. The review starts with a discussion of the issues identified through research to provide some background and context for the reset of the review. Research activities are grouped into categories for discussion. These categories are: building the infrastructure, pilot projects and digital preservation. The issues that are not currently being addressed are identified and conclusions are drawn. (Contains 51 references.) (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 825

Legal Deposit of Digital Publications: A Review of Research and Development Activity

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Cotton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Adrienne Muir
Department of Information Science
Loughborough University
Loughborough LE11 3TU, UK
+44 (0)1509 223065
A.Muir@lboro.ac.uk

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ABSTRACT

There is a global trend towards extending legal deposit to include digital publications in order to maintain comprehensive national archives. However, including digital publications in legal deposit regulation is not enough to ensure the long-term preservation of these publications. Concepts, principles and practices accepted and understood in the print environment, may have new meanings or no longer be appropriate in a networked environment. Mechanisms for identifying, selecting and depositing digital material either do not exist, or are inappropriate, for some kinds of digital publication. Work on developing digital preservation strategies is at an early stage. National and other deposit libraries are at the forefront of research and develop in this area, often working in partnership with other libraries, publishers and technology vendors. Most work is of a technical nature. There is some work on developing policies and strategies for managing digital resources. However, not all management issues or users needs are being addressed.

Categories and Subject Descriptors

K.5.0 **[Legal aspects of computing]**

General Terms

Management, Legal Aspects

Keywords

Legal deposit Digital publications Digital preservation

1. INTRODUCTION

The concept and practice of legal deposit is under threat in the digital environment. The main, though not the original, aim of legal deposit is to ensure the preservation of a nation's intellectual and cultural heritage over time. Many countries are extending legal deposit regulations to cover digital publications in order to maintain comprehensive national archives. However, even countries that have been dealing with the legal deposit of digital publications for some time are still grappling with how to collect and manage this material effectively in the long term. Existing collection management principles and practice were not designed with digital information in mind. Online and networked publications pose particularly complex challenges.

Publishers typically have a legal obligation to deliver one or more copies of their publications to deposit libraries. The depositor is usually responsible for the cost of deposit. Legal depositories include national libraries, parliamentary libraries, university libraries and national archives (for non-print material). There is great variety in the types of material collected through legal deposit. The requirement is usually material available to the public whether for sale, hire or for free. Printed publications are really the only common factor. Other types of material collected include sound recordings, audiovisual material and software. For a recent summary of the current status of legal deposit around the world see [1].

There is currently a great deal of research and development activity in this area. Early work focused on identifying issues and problems and on gathering information for making the case for extension of legal deposit. Currently deposit libraries are carrying out work, often in collaboration with other deposit libraries, publishers and technology vendors. Much of this work is of a technical nature and focuses on building the basic infrastructure, setting up digital depositories and collecting digital publications. Researchers are also working on metadata and digital preservation issues.

This review of research and development work will focus on activities specifically related to digital legal deposit. However, it will also touch on more generic work that is especially relevant. The review starts with a discussion of the issues identified through research to provide some background and context for the rest of the review. Research activities are grouped into categories for discussion. These categories are: building the infrastructure; pilot projects and digital preservation. The issues that are not currently being addressed are identified and conclusions are drawn.

2. THE ISSUES

The deposit of digital publications raises legal, economic, technical and managerial/organisational issues at all stages of the legal deposit process. For the purposes of this review, these stages are summarised as: identification of publications; selection; acquisition; accession and processing, including storing; preservation; and access. There are a number of fundamental factors that facilitate the legal deposit of digital publications: definitions; metadata; and standards.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
JCDL '01, June 24-28, 2001, Roanoke, Virginia, USA.
Copyright 2001 ACM 1-58113-345-6/01/0006…$5.00.

IR058361

BEST COPY AVAILABLE

There are also political issues associated with legal deposit. These arise because there are a number of different actors involved in the legal deposit system, and each of these actors has their own interests. The interests of one group do not necessarily coincide with another. For example, the commercial interests of publishers and the legal requirement to give up several copies of their product combine to cause tension between publishers, deposit libraries and legislators.

An important point that arises from the literature is that decisions taken at one stage affect decisions taken at other stages. Selection policies may need to take into account the ability of the depository to capture and preserve particular publications. Technical, legal, economic and organisational issues may influence preservation choices. Alternatively, different preservation strategies have different economic and management implications.

For the purposes of this review, the issues identified through research are discussed within the framework of the legal deposit process described above. Metadata and standards are also discussed in this way. The issue of definitions also pervades the whole process. Many well-established concepts either do not apply in the digital environment or need redefinition. This issue is so fundamental to legal deposit in the digital age that it is discussed separately.

3. DEFINITIONS

A common theme in the literature is a lack of agreed definitions for various concepts in the digital environment. Terms that are well understood in the print environment are irrelevant or have new meanings in the digital world. Examples include terms relating to documents or publishing such as "publication," "place of publication," "publication date," "publisher," "edition" and "authenticity." Another problem is that the same words have different meanings for different communities. For example, "archives" and "metadata" have different meanings for different professional communities [2]. Researchers working in this area have developed glossaries [3, 4]. Unfortunately, these are of limited use because they have been developed specifically for the use of project participants.

3.1 Documents, publications and publishing

Many of the problems of definition associated with legal deposit stem from the fact that the concept was originally based on mainly textual information first made available in individual nations via a physical carrier, usually a book. Some of the traditional concepts still apply to digital information on physical media. However, telecommunications and global networking have radically changed the nature of information dissemination. Many online publications are frequently, if not continuously, updated and they are globally available. New types of communication have emerged, such as email, mailing lists, chat rooms, personal World Wide Web home pages and dynamic Web pages that are generated 'on the fly' from databases. How much of this information can be called a "publication" in the traditional sense is open to conjecture.

The British Library commissioned a study on the definition of terms. Existing sources of definitions are given in appendices to the study report. The study found that, at the time, existing definitions were not helpful because they did not deal well with new types of material, including digital material [5]. One point made was that definitions should be format or medium-independent to make them "future proof" [5].

As far as the concepts of documents and publishing were concerned, the report did not suggest definitions, but provided "an overall framework of analysis within which to work" [5]. Martin defines a document as "(a) a combination of a work or compilation of works the medium on which the work or compilation is stored and any access technology which is specific to the document or (b) any one of a number of copies of such a combination."

This is a more complex definition than that of Mackenzie Owen and van der Walle. They define electronic publications as "published documents which are produced, distributed stored and used in electronic form" [6].

Martin also defines "published within the United Kingdom." This is "the public to which it is offered or broadcast or made available or before which it is performed includes a part of the United Kingdom … and the publisher or an importer or distributor or an agent of any of the aforementioned is domiciled in the United Kingdom" [5]. Martin admits that this definition creates a potential loophole for publishers whose entire operation is outside of the UK, but whose offering is directly at a UK audience. He also points out it would be difficult to enforce UK law in this situation. This problem would apply to any country.

Another British Library sponsored study [7] spells out the potential problems associated with depositing online material. The report provides definitions for different types of database. However, a major point made is that the traditional concept of publishing is not applicable in the digital environment. The "publication" process is not the same in the print and online environments and different entities are involved. In the online world, no single entity has overall control of the process and intellectual property rights are created at several points. The entity that owns the rights to the data may be different from the entity hosting it. The entity owning the rights to the retrieval software may be different from the data owner and/or the host entity. Who is responsible for deposit?

3.2 Preservation

There is also confusion in the terminology used for the preservation of digital information. For example, there is a difference between digital preservation and preservation digitisation [2]. Digital preservation is "the storage, maintenance, and accessibility of a digital object over time." Preservation digitisation involves digitising a fragile object to preserve its intellectual content. Preservation digitisation, in contrast, produces a surrogate for the original object. This surrogate will then need to be preserved over time.

4. IDENTIFICATION, SELECTION, ACQUISITION

4.1 Identification

The 1996 ELDEP study reported that the amount of material published only in digital form was quite small compared the volume of traditional publishing output [8]. This view was repeated in 1999 in another study [9]. However, both of these studies stated that the proportion of published output released only in digital form was likely to increase over time.

In order to acquire information, depositories have to identify it. One suggestion here is that legal deposit could require all publishers to register their publications [10]. The existence of publications would then be known, even they were not all collected. It may be

impossible to enforce in practice because of the large numbers involved, the ignorance of many Internet "publishers" of the traditional systems, or simply their unwillingness to comply.

Bibliographic control is well developed in the print environment, less well developed for non-print formats and, it seems, virtually non-existent for digital publications. One particular area of concern is that of unique identification of digital publications. While identification of offline digital material such as CD-ROMs may be reasonably straightforward using existing identifiers, online material presents problems [9]. There may be different manifestations of the same content. The question is whether each manifestation should have a different identifier, or whether there be one identifier for the underlying work. This raises the further question of how to identify each manifestation and relate this to the underlying work.

New types of identifier being developed include the Digital Object Identifier (DOI) and Uniform Resource Names. The DOI is being developed by the International Foundation to help in the management and exploitation of digital information [11]. Uniform Resource Names are persistent identifiers for online information [12].

4.2 Selection

Mackenzie Owen and de Walle point out that legal deposit laws are often selective in their coverage [6]. Some types of material are included and some are not. They recommend that, with some exceptions, all digital publications should be collected, including those published in parallel with print equivalents. An important point made is that deposit libraries will have to accept that they may never be able to collect all digital publications. There will be too many publications, too many publishers, and the rate of technological change is too fast.

There are some attempts at comprehensive collection of material. In Sweden, the National Library is attempting to capture the Swedish portion of the World Wide Web [13]. The aim of the Internet Archive is to archive the entire Internet [14]. The comprehensive approach may be feasible for countries with a relatively small digital publishing output, but it may turn out to be impossible for countries with bigger outputs.

Different depositories have different collecting policies, but selection often involves quality judgements: the importance of a particular publication, or its future research value. Technical issues can potentially distort digital selection policies [10]. It may well be that for crucially important material that is, for technical reasons, difficult but not impossible to acquire, access and preserve, expense is not a factor for consideration. There is the problem of moderately important "difficult" publications or crucially important publications that are impossible to deal with. These documents may end up being lost.

The acquisition of dynamic digital information is particularly problematic. Many writers comment on the impracticality or even impossibility of capturing every version of databases that are amended very frequently or in real-time [15]. The acquisition method here would be samples or snapshots. However, there is no commonly accepted practice and little practical experience of sampling techniques.

Hyperlinked documents present problems of deciding where the boundaries of the documents are. For example, there are questions about which is the appropriate level for archiving - a Web page, or an entire Web site just one big document? There is a question as to which linked sources should also be selected. Should only internal links within a document be maintained or should links to other documents be archived along with the original document?

With traditional publications, deposit usually means that some responsible entity sends physical objects to depositories. The situation is more complex in the digital environment. Online publications are not available in physical form so they cannot just be sent through the post. At present, there are three main options for acquiring online information. Publishers can transfer the information onto a physical medium and send that to the depositories. Publishers can arrange to transfer, or "push," information to depositories via networks. Alternatively, libraries can "pull" from publishers' sites themselves. A variant of this activity is "harvesting." This is usually done for Internet information, where the depositories use software to identify and pull in information from sites.

Several depositories are working with harvesting software to acquire publications, including the National Libraries of Australia [16], Finland [17] and Sweden [18]. Harvesting information is problematic in that existing tools do not entirely meet the needs of the depositories. Legal issues also arise, in that the depositories often need to negotiate permission from the publishers to copy their material. The push option may not work well on the Internet because it is populated with a huge number of publishers, some of whom are small organisations or even individuals. It would be impossible to set up relationships with all of them.

5. ACCESSION AND PROCESSING

Mackenzie Owen and Walle [8] recommend that quality checks and functional tests should be carried out for all items received. The purpose of such procedures is to check that the item is:

- The correct version
- In the required medium and format
- Complete
- Undamaged
- Error free and fully functional
- Not copy protected

Ensuring the authenticity of digital documents that are fluid by nature and capable of being changed very easily becomes a headache in the digital environment [15]. Some techniques for checking authenticity include time stamping and digital signatures.

The amount and types of information gathered at the accession stage will affect preservation of and long-term access to digital material. Deposit libraries will need more and better information, or metadata, from publishers from publishers at the point of accession than is necessary for printed information. There is also a need for some standardisation in this metadata.

The European Commission ELDEP study particularly focused on the bibliographic control of digital legal deposit collections [8]. There are questions as to whether current cataloguing rules can deal adequately with offline publications. The view emerging from the literature is that current rules may not be able to deal with online material at all.

6. PRESERVATION

Items in legal deposit collections are usually kept forever, therefore preservation is a central issue. If legal deposit collections are to include digital publications, solutions have to be found to the problems of digital preservation.

6.1 Media stability

Early preoccupations in this area were with the longevity of digital media. Estimates of the likely life expectancy of various storage media vary from around 1 to 100 years. Rothenberg gave some low estimates, including as little as two years for magnetic tape in some circumstances [19]. Unfortunately he gave no explanation for the low estimate and did not source his figures. The US National Media Laboratory contested Rothenberg's estimates [20]; it cites a 10-30 life expectancy for magnetic tape. Even so, this projection does not compare well with established archival media such as permanent paper or preservation microfilm. For these carriers, life expectancy is hundreds of years with optimal conditions.

As well as having inherent instabilities, the physical carriers used for digital information also react to environmental factors. These factors include both extremes of, and fluctuations in, temperature and relative humidity. Physical media also suffer from wear and tear and incorrect handling. Van Bogart produced a report on the storage and handling of magnetic tape, which is widely cited in the literature [21].

6.2 Technological obsolescence

Media instability is not the main problem as far as the preservation of digital information is concerned. The main problem is that viewing and using digital information requires the aid of equipment. The biggest threat to long-term survival is that of technological obsolescence of the hardware and software used to create and use digital information. Technical obsolescence is not a new problem for the preservation of information. Earlier examples of this include the Sony Betamax video recording format and Readex Microcards.

Recognition of technological obsolescence as the main threat to the long-term survival of digital information becomes prominent in the library and information science literature from the mid-1990s. Lehman sets out some of the aspects of technological change [22], including changes in coding and formats, software, operating systems and hardware. These changes can render digital material unreadable.

6.3 Preservation strategies

There are a number of possible strategies for digital preservation. A key question in deciding what strategy to use is what is to be preserved. Saving artefacts will not necessarily mean that the information itself is also preserved. Merely refreshing media will not overcome technological obsolescence. There are also problems associated with deciding exactly what the information or intellectual content is. This is especially problematic for multimedia or highly interactive information. Text, sound and pictures may be integrated; the software associated with the information may allow interaction between the user and the information. What has to be determined is whether it is the look and feel and functionality of the information product that is to be preserved, or just the raw information. [2].

There are a number of potential preservation strategies that address different preservation requirements and timeframes. The main preservation strategies are technology preservation, migration and emulation.

6.3.1 Technology preservation

Technology preservation is really a short-term strategy. This involves preserving the information in its original form and also the original software and hardware used to create and run the information. The strategy is likely to also involve media refreshment, especially for information stored on media with very short lifetimes [23]. However, hardware can only be maintained in working order for a finite period.

6.3.2 Migration

The Task Force on Archiving of Digital Information favoured the migration approach. The Task Force report defines migration as "the periodic transfer of digital material from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation." [24]. There are several migration strategies [25].

Adopting migration strategies means making long-term commitments to unknown future activities and unpredictable costs. Webb makes the point that most successful emulation work has been carried out with large amounts of homogenous data [10]. This is certainly not the situation for deposit libraries. While there have been some small-scale experiments with migrating publications from floppy disks [26], it is not known how well complex material stored on optical discs will migrate, if at all.

6.3.3 Emulation

The aim of the emulation concept is to allow long-term preservation of digital material while retaining the functionality and look and feel of the material. The main proponent of emulation as a preservation strategy for digital information is Jeff Rothenberg [27].

The idea of emulation is to view a digital document by using the software that created it. This does not necessarily mean that the software has to be run. The behaviour of the software could be described and the description saved so that its behaviour can be re-created in the future. The requirements for this approach would be to save the digital documents, the programs that were used to create the documents and all software required to run the documents. Software is dependent on the hardware it is created for, so the behaviour of an obsolete hardware platform would have to be emulated too. This would need the development of emulators, or software programs to mimic this behaviour [19].

Hardware emulation is potentially a simpler proposition than software emulation. The reasons for this is that there are fewer hardware platforms than operating systems and application software, so fewer emulators would have to be specified. Secondly, writing specifications for hardware is a better-developed practice than for software, so it would be easier to do [28].

In a paper for the Council on Library and Information Resources, Rothenberg set out the requirements for implementing emulation of hardware [28]. These include: techniques for specifying emulators; techniques for saving the necessary metadata (for finding, accessing and recreating documents) in human-readable form; and techniques for encapsulating documents, attendant metadata, software, and emulator specifications in a coherent and incorruptible way.

6.4 Disasters and rescue of digital information

Ross and Gow investigated approaches taken to access digital information when media are damaged or software and hardware are unavailable or unknown [29].

Things that can go wrong with digital information are:

- *Media degradation* – unfavourable environmental conditions in storage, disaster and manufacturer defects
- *Loss of functionality of access devices* – technological obsolescence, wear and tear of mechanical parts, lack of support for device drivers in newer software
- *Loss of manipulation capabilities* – due to changes in hardware and operating systems
- *Loss of presentation capabilities* - change in video display technologies, particular application packages may not run in newer environments
- *Weak links in the creation, storage and documentation chain* – data recovered but unreadable because of encoding strategy cannot be identified, loss of encryption documentation, use of unusual compression algorithms

Possible techniques for data recovery include heat and chemical treatments for soiled or damaged media, searching binary structures to identify recurring patterns and reverse engineering of content. One of the findings of the study is that there is a distinction between data recovery and data intelligibility. While it may be possible to recover data through searching binary structures, technological developments mean it will become harder to read the recovered data.

A future possibility is the use of magnetic force microscopy to read damaged media. Another is cryptography to help in the interpretation of recovered data. Ross and Gow also suggest an alternative to migration and emulation strategies for preservation. Retargetable binary translation involves "translating a binary executable programme from one machine … running a particular operating system … and using a particular file format … to another platform … running a different operation [sic] system … and using a different file format" [29].

6.5 Authenticity

Migration strategies potentially pose authenticity problems because they cause changes in the publications being migrated. Authenticity means that an object is the same as that expected based on a prior reference or that it is what it purports to be.

There is a range of strategies for asserting the authenticity of digital resources. The strategy used depends on the purpose for which authenticity is needed. These include unique document identifiers, the use of metadata to document changes, hashing, digital stamping, encapsulation techniques, digital watermarks and digital signatures.

During 1998, the CERBERUS project investigated the authenticity and integrity of electronic documents in digital libraries with a deposit task. The project partners were the Dutch Koninklijke Bibliotheek, the Technical Universities of Eindhoven and Delft and the University of Amsterdam. The project was co-funded by Innovation of Scientific Information Supply. There is a brief description on the KB Web site [30], but there is little written in English on the findings of this project.

6.6 Management of digital preservation: life cycles, stakeholders and rights issues

The concept of the life cycle of digital resources has been put forward as a tool for looking at the challenges of digital preservation. This was developed in the context of a study to develop a strategic policy framework for the creation, management and preservation of digital resources. Hendley added to the original model developed by the UK Arts and Humanities Data Service. The life cycle breaks down into several stages. These are: resource creation; selection and evaluation; management; disclosure; use; preservation; and rights management.

Data archives can often dictate requirements at the creation extent to a great extent. This is not the case for legal depositories. Deposit material and depositors will be diverse. In many cases, the main priority of depositors is commercial gain and this can conflict with preservation interests. The creators of the framework acknowledge this.

Rights issues are important because preservation strategies involve copying, and possibly changing, the original information in some way. The life cycle framework illustrates how the stages are interrelated and how decisions taken at one stage impact on other stages. The aim is to help in the policy and decision making process and help identify where collaboration efforts would help preservation. The framework is supported by a number of case studies. One case study is legal deposit libraries.

Haynes et al. examine the attitudes of "originators and rights holders" towards to the issue of their responsibility for digital preservation [31]. The study report lists various stakeholder groups. These are: libraries; publishers; archive centres; distributors; IT suppliers; legal depositories; consortia; authors; and networked information service providers.

The consultants made a number of recommendations in their report. One was that a body should be established to co-ordinate digital archiving activities - a National Office of Digital Archiving. This suggestion is similar to that of a national digital preservation officer from Matthews, Poulter and Blagg [32]. This idea has since been taken forward in the UK. The Joint Information Systems Committee of the Higher and Further Education Funding Councils has appointed an officer to develop digital preservation strategies and work with other bodies to establish a so-called Digital Preservation Coalition [33].

Another suggestion was for a distributed approach to digital preservation. Distribution could have a number of bases, including regional, format and ownership. The suggested National Office of Digital Archiving would coordinate the development of standards and guidelines in cooperation with other agencies. The consultants suggest that legal deposit should be used as a mechanism for acquiring material, but that publishers should only have to contribute one copy of each publication.

Users should not be charged for access, but costs should be shared between research funders, the public (through government funding), and research communities. There is no mention of any responsibility falling on publishers for maintaining archives.

More recently, researchers from AHDS have carried out a study on the preservation management of digital materials and have produced a draft workbook for preservation managers. The workbook brings together research findings and available guidelines and augments this through some original research with some case study organisations [34].

NORDINFO supported a study on the copyright questions related to the legal deposit of online material. This study concentrated on the European and Nordic legal environments. The study report concludes that there is a gap between copyright provisions and legal deposit objectives. Digital preservation requires copying, and copyright exceptions should allow this. Another point that comes up is that there may be moral rights issues arising from migration activities if they result in changes to the migrated material. The report suggests that technology, including electronic copyright management systems, can contribute to solving problems. There also needs to be some investigation into how depositories can cooperate with publishers to solve problems [35].

6.7 Costs

There is a serious problem with identifying the costs associated with digital preservation. Until full-scale operational systems have been running for some time, the nature and extent of costs cannot be known for certain. Some organisations have estimated the cost of caring for digital information. The British Library included some alternative costings in its proposal for the extension of legal deposit in the UK [36].

Hendley's study of different preservation methods and associated costs took into account diversity in digital materials [23]. The study drew on the work of a related set of studies, in particular the work carried out by the AHDS. However, it also reviewed other cost models and visited digital libraries and archives.

7. ACCESS

Traditionally, deposit libraries provide access to deposit publications without charge. It is clear that this situation may not be accepted by rightsholders in the digital environment [37]. It is likely that access to such deposited digital publications will be governed by licence agreements. Williamson gives a flavour of the potential complexity of providing access to digital information [38].

There is nothing in the literature that directly reports the views of users, so it is not clear what kind of access they would require. Access to printed legal deposit publications is often on a reference only basis. It is not clear whether users would be happy with the digital equivalent of this, or if they would want remote access. Neither is there any evidence that users would object to limitations on access rights, say limited or no printing or downloading of material.

Another legal issue is that of security. Authentication of users and the setting up of access rights are security measures that have implications for users and libraries. This issue is applicable to all types of digital library.

New technology provides the means of closely monitoring information use. The purpose of monitoring use may be to police user behaviour to ensure access agreements with publishers are not breached. Williams points out that logging use may be burdensome for deposit libraries [38]. There is also the potential for other uses of data, for example providing feedback to publishers, which may have data protection implications.

While stating that libraries accept that "the legitimate interests of publishers require that access is limited and controlled," Mackenzie Owen and Walle suggest that access should be encouraged to facilitate preservation of digital publications [8]. Their reason is that if electronic publications are not used for some time, they may be found to no longer work. On the other hand, a high level of access will "check the operability of electronic publications and … identify and remedy any access problems that may occur."

Before using digital information, users have to be able to find it. In the digital environment, bibliographic records can have direct pointers to the material. However, as Mackenzie Owen and Walle point out, there can be several pointers, including to the original storage and the archival storage location [8]. Both of these may be physically located in the library. In the case of online publications, the original location may be a network address.

8. BUILDING THE INFRASTRUCTURE

8.1 Open Archival Information System reference model

A major initiative relevant to digital legal deposit collections is the development of the Open Archival Information System (OAIS) Reference Model [39]. The Consultative Committee on Space Data Systems is drafting this standard for the International Standards Organisation. An OAIS archive preserves information for access and use by a so-called Designated Community. This model is not just applicable to an organization that stores digital records; it can be applied to any type of digital paper material. The OAIS models the functions involved in the long-term storage of and access to digital information. These functions include acquisition and processing (ingest), archival storage, access, data management and administration of the archive.

The development of the OAIS has influenced other work being carried out in exploring the development of digital deposit collections. The CEDARS (CURL Exemplars for Digital Archives) project in the UK has used OAIS in developing its preservation metadata specification. The European NEDLIB (Networked European Deposit Library) project is working on developing an infrastructure for a European digital deposit collection [40]. British Library and the Koninklijke Bibliotheek in the Netherlands have used OAIS as the basis of their recent tenders for systems to manage their digital collections.

8.2 The NEDLIB Project

The NEDLIB project started at the beginning of 1998 and finishes at the end of 2000. Funding comes from European Commission, and the project leader is the Koninklike Bibliotheek in the Netherlands. The project partners are eight European national libraries, one national archive, three publishers and two information and communication technology companies. This project should be very influential in helping depositories cope with digital material.

The stated project aim [40] is to "develop a common architectural framework and basic tools for building deposit systems for electronic publications." The project deals with the technical issues involved in extending legal deposit to digital material. A great deal of detailed project material is available on the project Web site [41].

The project consortium adopted the OAIS model, but the NEDLIB Deposit System for Electronic Publications (DSEP) will be narrower in scope than the OAIS model. This is because some of the OAIS functions, such as Data Management and Access, are part of the general digital library environment and not specific to the digital depository. The aim is to link the functions of the deposit system and the digital library environment through interfaces.

NEDLIB is working with Jeff Rothenberg on an emulation experiment. The plan is that the first stage will result in a design for the whole experiment, a plan for testing and comparing the results of the emulations with the original works and a framework of preservation criteria and authenticity characteristics. The second stage involves modelling the emulation process and identifying metadata and functionality requirements. The last stage of the emulation experiment will be the implementation and evaluation of the emulation process in the testbed developed by the NEDLIB project

The first stage of the emulation experiment is now complete. Rothenberg concludes that "The results of this study suggest that using software emulation to reproduce the behaviour of obsolete computing platforms on newer platforms offers a way of running a digital document's original software in the far future, thereby recreating the content, behaviour, and 'look-and-feel' of the original document" [42]. This claim seems somewhat inflated since the actual experiment actually involved running Windows 95 publications on an Apple Mac using Connectix VirtualPC software as the emulator. The most Rothenberg can claim is that this particular software does what it says it does.

8.3 The BIBLINK Project

BIBLINK started in 1996 with support from the European Commission. Although the project came to an end in 1999, work on the initiative is ongoing under the aegis of the Conference of Directors of National Libraries.

The original aim of BIBLINK was to help develop and improve national bibliographic services, focusing on digital publications, especially online publications. Potentially all libraries would be beneficiaries of the project. However, the main perceived benefit was that national libraries would not miss the publication of significant publications [43].

The BIBLINK project developed a prototype demonstration system, called the BIBLINK Workspace. The demonstrator provides a virtual workspace or "computer mediated work environment" for participating parties. It allows publishers to create records and allows participants to access the system to retrieve, update and delete records in the workspace. BIBLINK is developing an Exploitation Plan [43]. This will provide a framework for library partners to assess the possibility of incorporating the system into operational procedures.

8.4 CURL Exemplars for Digital Archives (CEDARS) Project

The CEDARS project in the UK is funded by the Joint Information Systems Committee of the Higher and Further Education Funding Councils through the eLib Programme. CEDARS began in April 1998 and is due to finish in March 2001. The aim of CEDARS is to 'address strategic, methodological and practical issues and provide guidance in best practice for digital preservation' [44].

The CEDARS team is a partner in a new project on emulation for preservation funded through the JISC/NSF (US National Science Foundation) International Digital Libraries Programme [45]. The other project partners are based in the University of Michigan. The project will 'develop a small suite of emulation tools, evaluate the costs and benefits of emulation as a preservation strategy for complex multi-media documents and objects, and develop models for collection management decisions that would assist people in making 'real life' decisions.

9. PILOT DEPOSITORIES

9.1 National Library of Canada

The National Library of Canada ran such a project between 1994 and 1995. The purpose of the Electronic Publications Pilot Project (EPPP) was to pilot the acquisition, cataloguing, preservation and provision of access to a few Canadian electronic journals and other publications available via the Internet [46]. The National Library of Canada is now building a full-scale electronic collection.

9.2 Koninklijke Bibliotheek, Netherlands

The Koninklijke Bibliotheek in the Netherlands took the decision to collect digital publications in 1994 [47, 48]. Offline publications, such as CD-ROMs were stored on the stacks with the books. From 1995, the KB experimented in handling online publications on a small scale. Three publishers cooperated with the KB by agreeing to deposit some of their electronic publications with the KB. In 1996, the KB reached a provisional agreement with publishers to widen deposit. This small-scale deposit system was based on the IBM Digital Library system and became operational in 1998 [48]. The KB is now setting up a full-scale system [49].

9.3 National Library of Australia

The National Library of Australia set up the PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) project in 1996. The aim of the project was to "develop policies and procedures for the selection, capture, archiving and provision of long-term access to Australian electronic publications." The Library developed a proof-of-concept archive of Australian Internet material, which has been used to develop policies and procedures for the long-term preservation and access to digital publications [50].

The National Library of Australia realised that it needed integrated systems for managing all parts of its collections, including digital material. The NLA is taking this forward with its Digital Services Project. This project will provide storage for its digital material, but it will also provide management systems for most of the Library's collections.

9.4 Helsinki University Library (National Library of Finland)

EVA was originally an eighteen month Finnish project that started in June 1997. The aim of the project was "to test methods of capturing, registration, preserving and providing access to … online documents …" [51]. EVA was used to test tools being developed by various Nordic projects, including a Dublin Core metadata template and converter, URN generator and a harvesting and indexing application. Documents were harvested from the World Wide Web using the harvester. Once captured, the documents were analysed,

indexed then archived. The EVA II and EVA III projects have been building on this work [52].

9.5 Kungliga Biblioteket, Sweden

The Kungliga Biblioteket in Sweden is currently running the Kulturarw3 project. The aim of this project is to "to test methods of collecting, preserving and providing access to Swedish electronic documents which are accessible on line in such a way that they can be regarded as published." [18]. The project aims to collect Swedish material available on the Internet according to specified selection criteria and to automate the collection through the use of robots.

10. CONCLUSIONS

There is a great deal of activity in this area worldwide. While exploratory work has identified many problems arising from the legal deposit of digital publications, a great deal more work is needed to solve these problems. Much of the work has been small-scale or national, yet the problems transcend national boundaries. Therefore, initiatives such as NEDLIB are to be welcomed.

The OAIS reference model provides a conceptual outline of the processes involved in a digital depository. However, by its nature it does not consider how theseprocess will be carried out. The NEDLIB project is attempting to develop a generic technical infrastructure for digital depositories. However, the project is limited in its scope and does not deal with the interface between the depository and the depositors.

Much of the activity in this area is concerned with technical issues. There is little evidence of any work taking a wide view of managerial or organisational issues. These issues include the management of workflows in depositories, staffing and skills requirements. It is also likely that there will have to be a lot more cooperation between publishers and depositories to facilitate deposit, preservation and access. Access especially may require negotiation. While there is an increasing interest in user needs in digital library research, there is little evidence of this in the legal deposit context.

Current work assumes that deposit systems will be organised in a similar way to current systems in that material will be physically deposited in deposit libraries. Physical deposit may be necessary to ensure long-term preservation, but alternatives to the current system could be considered. While the concept of a comprehensive archive of the national intellectual output may remain an ideal in an increasingly knowledge intensive world, it is not yet clearwhether this is technically and organisationally feasible or affordable.

11. ACKNOWLEDGEMENTS

I wish to acknowledge part-funding of this research by the Loughborough University Faculty Studentship scheme.

12. REFERENCES

[1] Muir, A. Legal deposit of digital material in the UK: recent developments and the international context. Alexandria, 12(3), (2000), 151-165.

[2] Russell, K. Digital preservation: ensuring access to digital materials into the future. http://www.leeds.ac.uk/cedars/Chapter.html.

[3] Working Definitions of Commonly Used Terms (for the purposes of the Cedars Project). http://www.leeds.ac.uk/cedars/documents/PSW01.htm.

[4] Borbinha, J., Cardoso, F., and Freire, N., NEDLIB glossary. http://www.kb.nl/coop/nedlib/glossary.pdf.

[5] Martin, D. Definitions of publications and associated terms in electronic publications. British Library Research & Development Department, 1996.

[6] MacKenzie, G. Searching for solutions: electronic records problems worldwide. Managing information, (July/August 2000), 59-65.

[7] EPS Ltd. The legal deposit of online databases. British Library Research & Development Department, 1996.

[8] Mackenzie Owen, J.S.and Walle, J.v.d. Deposit collections of electronic publications. Office for Official Publications of the European Communities, 1996.

[9] Bide, M., Potter, E.J., and Watkinson, A. Digital preservation: an introduction to the standards issues surrounding the archiving of non-print material. Book Industry Communication, 1999.

[10] Webb, C. Long-term management and preservation of publications on CD-ROMs and floppy disks: technical issues. http://www.nal.gov.au/niac/meetings/tech.html.

[11] The Digital Object Identifier System. http://www.doi.org/.

[12] Uniform Resource Names (urn). http://www.ietf.org/html.charters/urn-charter.html.

[13] Arvidson, A. and Lettenstrom, F. The Kulturarw3 Project - the Swedish Royal Web Archive. The Electronic Library, 16(2 April 1998), 105-108.

[14] Kahle, B. Archiving the Internet. http://www.archive.org/sciam_article.html.

[15] Mandel, C.A. Enduring access to digital information: understanding the challenge. LIBER quarterly, 6 (1996), 453-464.

[16] Digital Services Project. http://www.nla.gov.au/dsp/.

[17] Lounamaa, K. and Salonharju, I. EVA- the acquisition and archiving of electronic network publications in Finland. Tietolinja news, 1 (1999). http://www.lib.helsinki.fi/tietolinja/0199/evaart.html.

[18] Kulturarw3 heritage project. http://kulturarw3.kb.se/html/projectdescription.html.

[19] Rothenberg, J. Ensuring the longevity of digital information. http://www.clir.org/programs/otheractiv/ensuring.pdf.

[20] Van Bogart, J.W.C. Mag tape life expectancy 10-30 years. http://palimpsest.stanford.edu/bytopic/electronic-records/electronic-storage-media/bogart.html.

[21] Van Bogart, J.W.C. Magnetic tape storage and handling. http://www.clir.org/pubs/reports/pub54/.

[22] Lehmann, K.-D..Making the transitory permanent: the intellectual heritage in a digitized world of knowledge in Books, bricks & bytes. American Academy of Arts and Sciences, Cambridge, Mass., 1996, 307-329.

[23] Hendley, T. Comparison and methods & costs of digital preservation. British Library Research and Innovation Centre, 1998.

[24] Task Force on Archiving of Digital Information. Preserving digital information: report of the Task Force on Archiving of Digital Information. RLG, CPA, 1996.

[25] Feeney, M. Towards a national strategy for archiving digital materials. Alexandria, 11(2), (1999), 107-121.

[26] Woodyard, D. Farewell my floppy: a strategy for migration of digital information. http://www.nla.gov.au/nla/staffpaper/valadw.html.

[27] Rothenberg, J. Ensuring the longevity of digital documents. Scientific American, 272(1), (1995), 42-47.

[28] Rothenberg, J. Avoiding technological quicksand: finding a viable technical foundation for digital preservation. Council on Library and Information Resources, 1999.

[29] Ross, S.and Gow, A., Digital archaeology: the recovery of digital materials at risk. British Library Research and Innovation Centre, 1999.

[30] CERBERUS. http://www.kb.nl/kb/sbo/dnep/cerberus-en.html.

[31] Haynes, D., et al. Responsibility for digital archiving and long term access to digital data. Library Information Technology Centre, 1997.

[32] Matthews, G., Poulter, A., and Blagg, E., Preservation of digital materials policy and strategy issues for the UK. British Library Research and Innovation Centre, 1996.

[33] Digital Preservation. http://www.jisc.ac.uk/dner/preservation/.

[34] [34] Jones, M.and Beagrie, N. Preservation management of digital materials workbook: pre-publication draft, . 2000.

[35] Mauritzen, I.and Solbakk, S.A. A Study on copyright and legal deposit of online documents. NORDINFO, 2000.

[36] British Library, Proposal for the legal deposit of non-print publications to the Department of National Heritage from the British Library. British Library, 1996.

[37] Wille, N.E. Legal deposit of electronic publications: questions of scope and criteria for selection in Legal deposit with special reference to the archiving of Electronic publications: proceedings of a seminar organised by NORDINFO and the British Library (Research and Development Department) (Windsor, England 27-29 October 1994). NORDINFO, 1994.

[38] Williamson, R. Access and security issues from a publisher's perspective in Legal deposit with special reference to the archiving of Electronic publications: proceedings of a seminar organised by NORDINFO and the British Library (Research and Development Department) (Windsor, England 27-29 October 1994). NORDINFO, 1994.

[39] Consultative Committee for Space Data Systems. Reference model for an Open Archival Information System (OAIS). http://ftp.ccsds.org/ccsds/documents/pdf/CCSDS-650.0-R-1.pdf.

[40] Werf-Davelaar, T. NEDLIB: Networked European deposit library. Exploit interactive, 4 (2000). http://www.exploit-lib.org/issue4/nedlib/.

[41] NEDLIB. http://www.kb.nl/coop/nedlib/homeflash.html.

[42] Rothenberg, J. An Experiment in using emulation to preserve digital publications. http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf.

[43] Noordemeer, T.C. A Bibliographic link between publishers of electronic resources and national bibliographic agencies: Project BIBLINK. Exploit interactive, 4, (2000). http://www.exploit-lib.org/issue4/biblink.Cedars project summary. http://www.leeds.ac.uk/cedars/documents/MGA04.htm.

[44] JISC/NSF International Digital Libraries Programme: Emulation for Preservation: Project Summary. http://www.leeds.ac.uk/cedars/JISCNSF/summary.htm.

[45] Electronic Publications Pilot Project Team. Electronic Publications Pilot Project (EPPP): final report. http://www.nlc-bnc.ca/pubs/abs/eppp/ereport.htm

[46] Noordemeer, T.C., Deposit for Dutch electronic publications: research and practice in the Netherlands in Research and advanced technology for digital libraries: first European Conference, ECDL'97 (Pisa, Italy, September 1-3, 1997). Springer, 1997.

[47] Steenbakkers, J. Developing the Depository of Netherlands Electronic Publications. Alexandria, 11(2), (1999), 93-105.

[48] Werf-Davelaar, T.v.d. DNEP 1995-2000: the Dutch Deposit of Electronic Publications. http://www.bic.org.uk/Titia.ppt.

[49] Relf, F.A. PANDORA - towards a national collection of Australian electronic publications. http://www.nla.gov.au/nla/staffpaper/ashrelf1.html#abst.

[50] Eva - the Acquisitions and archiving of electronic network publications. http://renki.lib.helsinki.fi/eva/english.html.

[51] Eva - elektronisen verkkoaineiston hankinta ja arkistointi. http://www.lib.helsinki.fi/eva/index.html.